Motif analysis of amphioxus, lamprey and invertebrate estrogen receptors and amphioxus and human estrogen-related receptors: Towards a better understanding of estrogen receptor evolution

*Michael E. Baker[1] and Charlie Chandsawangbhuwana[2]

[1]Department of Medicine, 0693

[2]Department of Bioengineering

University of California, San Diego

9500 Gilman Drive

La Jolla, CA 92093-0693

*Corresponding author

E-mail addresses:

MEB: mbaker@ucsd.edu

CC: cchandsa@ucsd.edu

Abstract

Background. The origins of steroid-dependent regulation of the vertebrate estrogen receptor (ER) are poorly understood. Genes with statistically significant sequence similarity to vertebrate ERs have been found in lamprey, a basal vertebrate, and amphioxus, a basal chordate. Motif analysis of these sequences provides an opportunity to investigate early events in the evolution of the ER.

Results. We used artificial intelligence-based software to construct twelve motifs specific to the estrogen-binding domain of ERα and ERβ in land vertebrates and teleosts. We mapped these ER-specific motifs onto the sequences of lamprey, amphioxus, invertebrate and selected vertebrate ERs and amphioxus and human estrogen-related receptor (ERR). We find that lamprey ER contains eleven motifs common to ERs in the training set. In contrast, amphioxus ER contains only six motifs. Various invertebrate ERs contain either seven or eight motifs. Unexpectedly, human and amphioxus ERRs contain nine of the twelve motifs, despite extensive sequence divergence during the descent of chordate ERs and ERRs from a common ancestor. We mapped the twelve motifs onto a multiple alignment of human, lamprey and amphioxus ERs, which depicted residues in human ERα that are known to bind estradiol. There is excellent

conservation of these key residues in lamprey ER and poor conservation in amphioxus ER. Out of seventeen residues on human ERα that bind estradiol, sixteen and six are identical in lamprey and amphioxus ER, respectively. A phylogenetic tree of ERs and ERRs reveals a long branch for amphioxus ER, which is in agreement with the low sequence and motif similarity between amphioxus ER and other ERs.

Conclusions. There are significant differences between *B. floridae* ER and vertebrate ERs in the steroid-binding domain as measured by motif analysis and percent of amino acids that are known to stabilize estradiol in human ERα. This suggests that novel steroids regulate transcriptional activity of *B. floridae* ER. The absence in lamprey ER of motif 10, which maps to the c-terminus half of α-helix 9, may be important in recognition of novel estrogens, such as 15α-hydroxy-estradiol.

Key words: estrogen receptor evolution, amphioxus estrogen receptor, lamprey estrogen receptor, invertebrate estrogen receptor, estrogen-related receptor

Background

The physiological actions of estradiol are mediated by binding to the estrogen receptor [ER], which belongs to the nuclear receptor family, a large and diverse family of transcription factors [1-5]. Other steroid receptors in this family include the androgen receptor (AR), glucocorticoid receptor (GR), mineralocorticoid receptor (MR) and progesterone receptor (PR). Phylogenetic analyses of the sequences of their steroid-binding domain reveal that AR, PR, GR and MR cluster together, while the ER is on a separate branch [1, 2, 4, 5]

Insights into when adrenal and sex steroids began to regulate gene transcription have come from sequencing of genomes of bacteria, yeast and plants. These genomes do not contain either steroid receptors or other nuclear receptors, indicating that nuclear receptors arose in multicellular animals [2, 4-6]. Other analyses revealed that adrenal and sex steroid receptors are absent from the fruit fly and roundworm, although these two invertebrates have nuclear receptors. Recently, the sea urchin genome, a basal deuterostome [Figure 1], was sequenced and found to contain 33 nuclear receptors [7]. None of these nuclear receptors, however, was a steroid receptor. *Ciona*, a chordate close to vertebrates, also does not contain steroid receptors [8].

Although, distant relatives of the ER have been found in octopus, snails and other mollusks, these ERs are active in the absence of steroids and do not bind estradiol or other steroids [9-13].

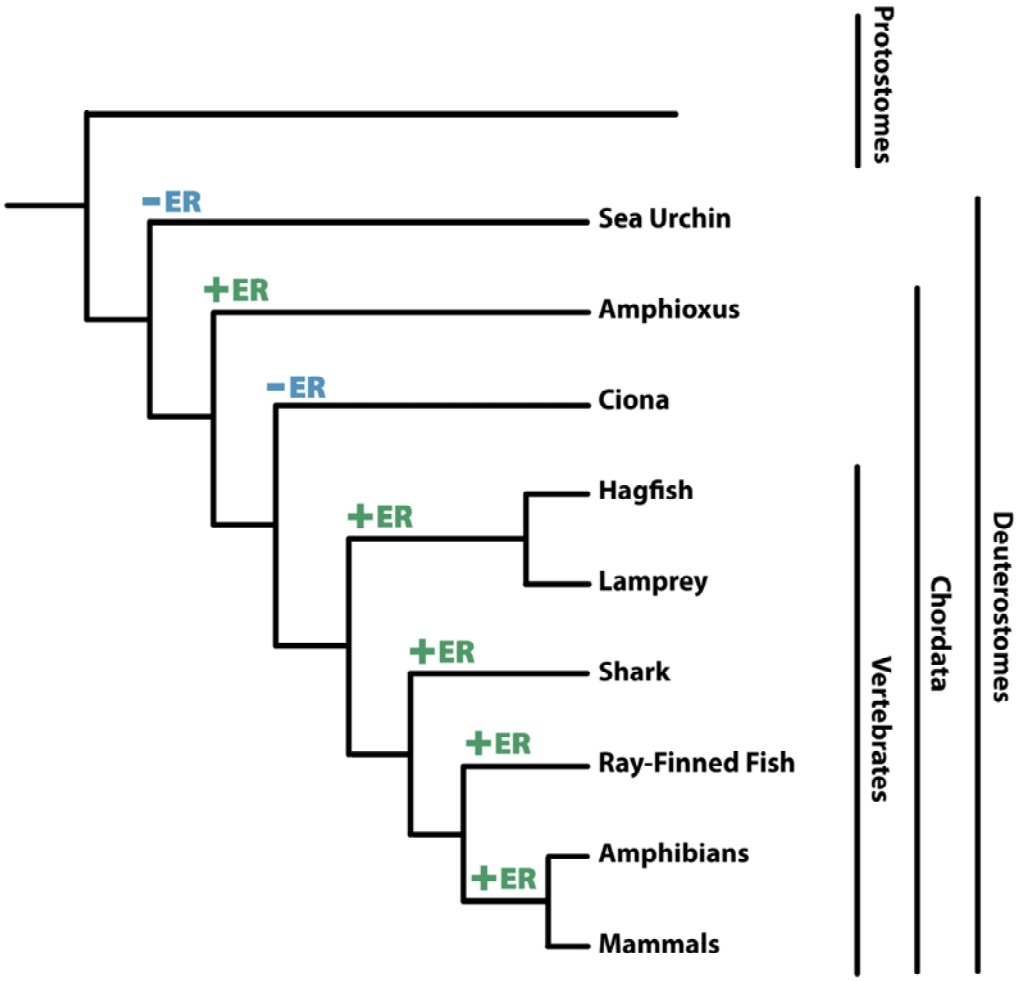

Figure 1. Phylogeny of amphioxus and other aeuterostomes.
Sea urchins are at the base of the deuterostome line. Amphioxus and *Ciona* are close relatives to vertebrates. *Ciona* [8] and sea urchin [7] do not contain steroid receptors; hagfish [14] and lamprey [15] contain steroid receptors. Distant relatives of the ER have been found in invertebrates, but these ERs do not bind estradiol or other steroids [9-13].

Thus, at this time, the most primitive animals that contain clear orthologs of adrenal and sex steroid receptors are lamprey and hagfish [14, 15], which are jawless fish at the base of the vertebrate line [Figure 1]. Recently, evidence from two sources supports the presence of an estrogen receptor in amphioxus, a basal chordate. First is the report that *Branchiostoma belcheri*, a pacific amphioxus, contains cytochrome P450s that are necessary for the synthesis of the sex

steroids estradiol, testosterone and progesterone, and that these steroids are present in reproductive organs in *B. belcheri* [16]. Second is the cloning of a gene from *B. floridae* that has statistically strong sequence similarity to vertebrate ERs [Genbank:ABQ42696] by Ciaccia *et al*. at Boston University and by the Joint Genome Initiative [JGI:210589]. Data on the binding of estradiol or other steroids to this amphioxus protein has not been reported.

Our interest in the origin and evolution of adrenal and sex steroid action [5, 6, 17, 18] motivated us to investigate the relationship of the amphioxus and lamprey ERs to teleost and land vertebrate ERs. With this goal in mind, we used artificial intelligence-based software [19, 20] to construct twelve motifs specific to the estrogen-binding domain of ERα and ERβ in teleosts and land vertebrates, and mapped these ER-specific motifs onto amphioxus ER and lamprey ER and used motif analysis to investigate the relationship of invertebrate ERs and the ERR [21-24] to estrogen-binding ERs. We find that lamprey ER contains eleven of these twelve motifs, while amphioxus ER contains only six motifs. We also find that various invertebrate ERs contain either seven or eight of these motifs. Unexpectedly, amphioxus and human ERRs contain nine motifs. It appears that despite extensive sequence divergence during the descent of ER and ERRs from a common ancestor [22-24] several ancestral motifs were conserved.

Results

Motif analysis of ERα and ERβ in land vertebrates and teleosts

MEME determined twelve motifs from three training sets: ERα alone, ERβ alone and the combined ERα and ERβ set, which were mapped by MAST onto various ERs and ERRs in schematics shown in Figure 2 [ERα alone], Figure 3 [ERβ alone] and Figure 4 [combined ERα and ERβ set].

ERα Training Set

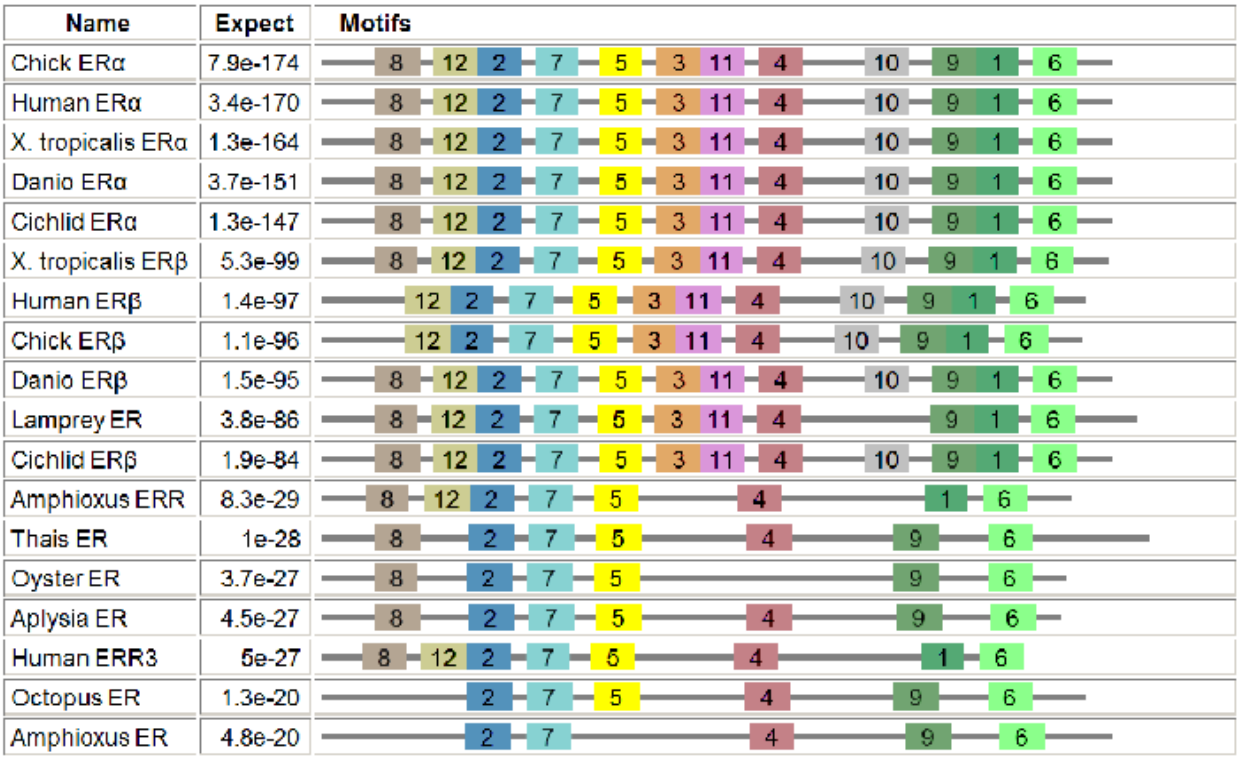

Figure 2. MAST analysis of MEME motifs for the ERα training set
Twelve motifs calculated by MEME for the ERα training set were mapped by MAST onto the sequences of various ERs and ERRs. Comparisons with the data in Figures 3 and 4 reveals that the position of some motifs differs for human ERα and ERβ, depending on the training set. The E-value of a sequence in the MAST output is the expected number of sequences in a random database that would match the motifs as well as the sequence does.

ERβ Training Set

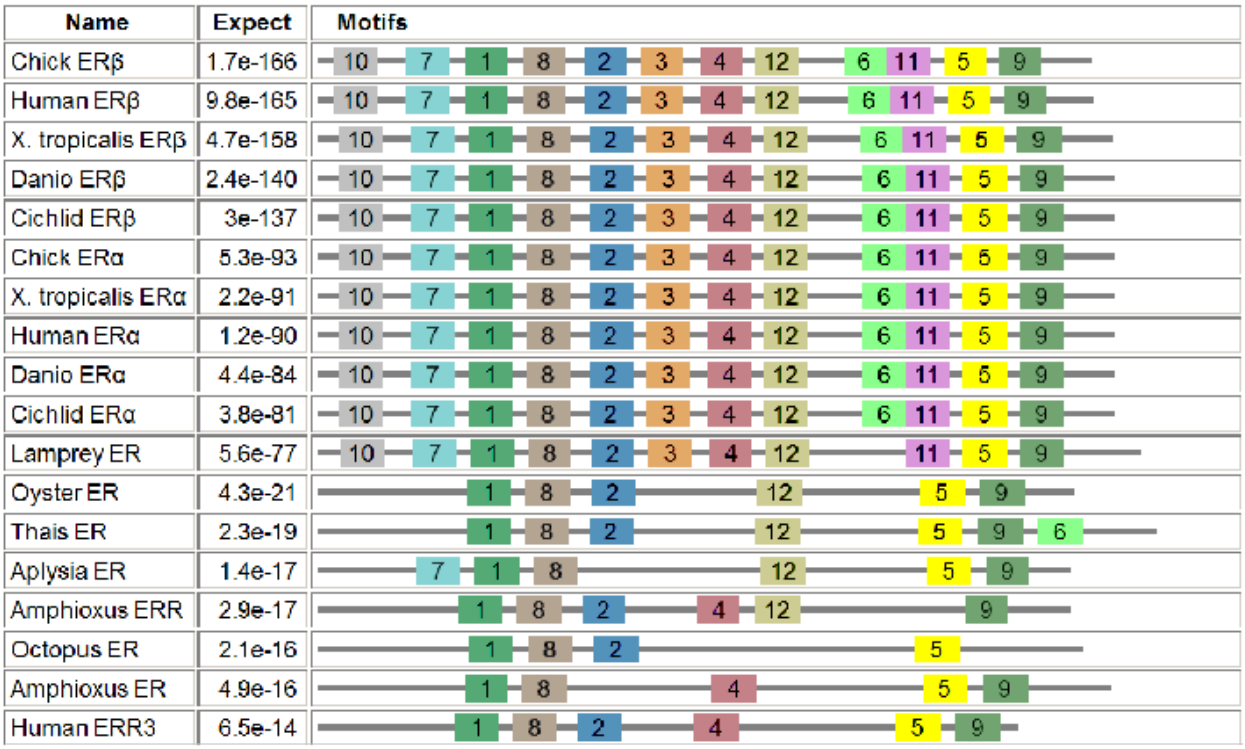

Figure 3. MAST analysis of MEME motifs for ERβ training set
Twelve motifs calculated by MEME for the ERβ training set were mapped by MAST onto the sequences of various ERs and ERRs.

ERα & ERβ Training Set

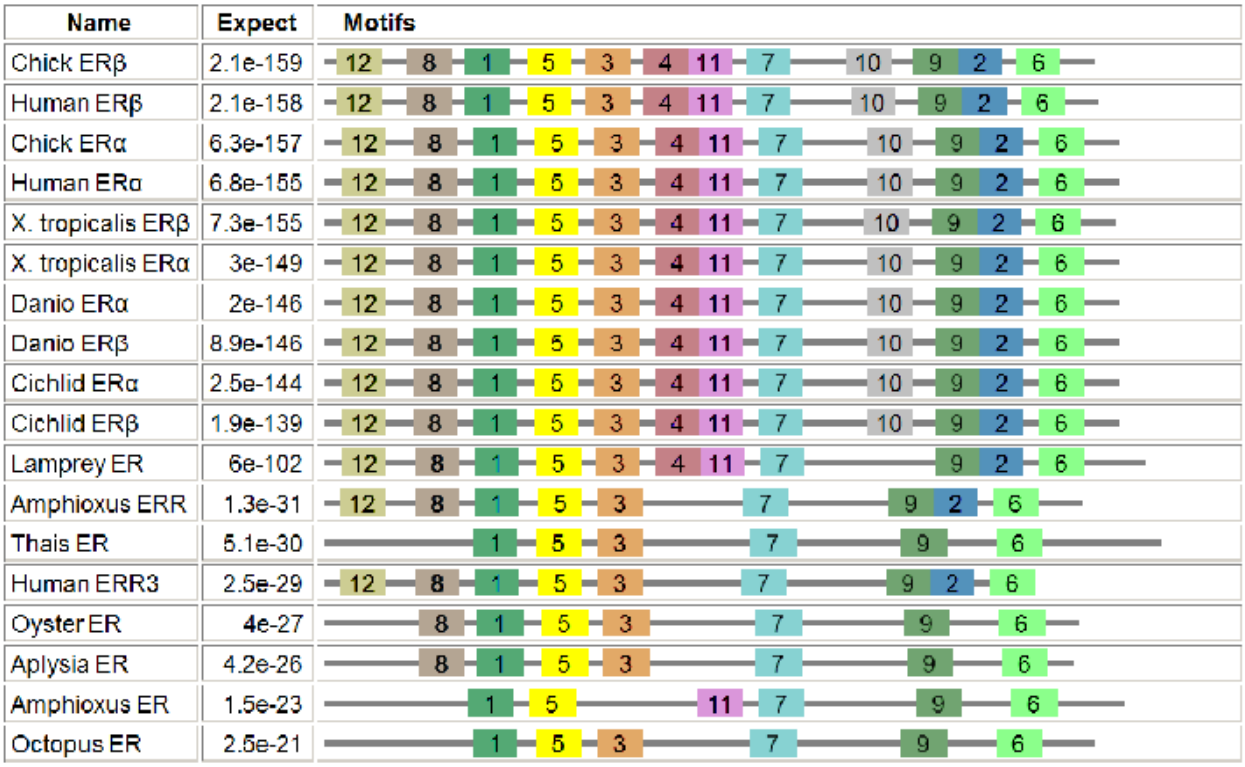

Figure 4. MAST analysis of MEME motifs for the combined ERα and ERβ training set
Twelve motifs calculated by MEME for the combined ERα and ERβ training set were mapped by MAST onto the sequences of various ERs and ERRs. Lamprey lacks motif 10. Amphioxus ER contains six motifs. Amphioxus and human ERR contain nine motifs. Invertebrate ERs contain seven or eight motifs.

MAST also calculated a statistical score for the motifs for each ER. As expected, lamprey ER is closest to the other ERs [Figures 2-4]. Unexpectedly, the MAST score for amphioxus ERR is closer to vertebrate ERs than is amphioxus ER. Various invertebrate ERs and human ERR have more distant MAST scores.

Although all twelve motifs are found in the land vertebrate and teleost ERs, there are differences in the location of the motifs, which depend on the training set used by MEME [Figures 2-5]. For example, motif 1 for the ERα training set [Figure 2] and ERβ training set [Figure 3] maps to different parts of these receptors. Thus, motif analysis distinguishes between ERα and ERβ. The motifs for the combined ERα and ERβ set [Figure 4] have some similarities and differences with motif patterns in the two individual training sets. Because the combined ERα and ERβ training set contains the most information about the estrogen-binding domain of

ERs, we will use MEME and MAST results with the combined training set for analysis of conservation of motifs in various ERs and ERRs.

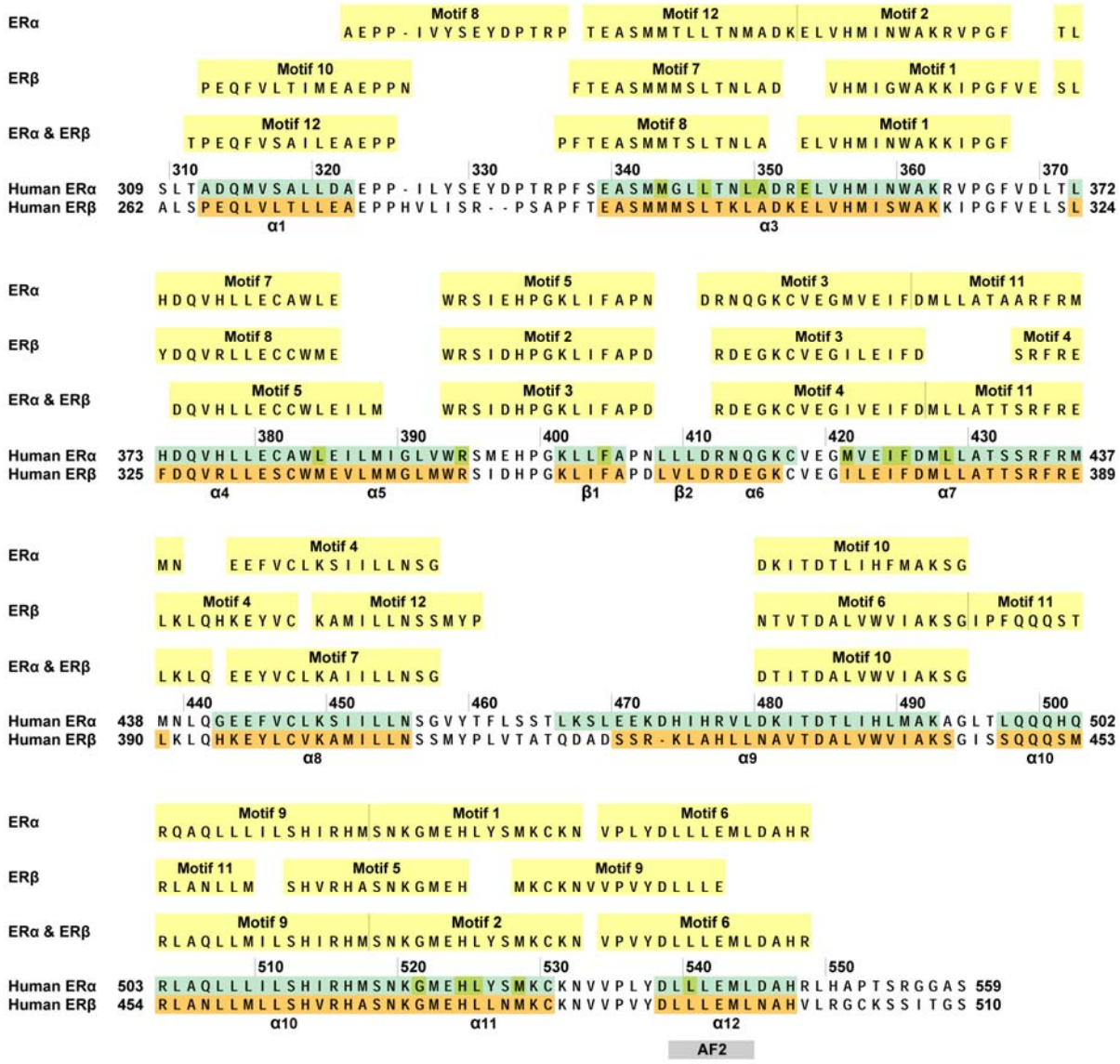

Figure 5. Mapping by MAST of MEME motifs from three training sets onto ER sequences. The twelve motifs calculated by MEME for the three training sets were mapped by MAST onto the sequences of human ERα and ERβ. The α-helices and β-strands from the crystal structures of human ERα and ERβ are shaded. Residues in human ERα that are involved in binding of estradiol [25] are shown in green.

Examination of Figure 4 reveals that lamprey contains eleven out of twelve motifs constructed from the combined ERα and ERβ training set. Amphioxus ER contains six of the

twelve motifs in this training set. Invertebrate ERs contain either seven or eight motifs, while amphioxus and human ERR contain nine of the twelve motifs.

Mapping of motifs on sequences of human ERα, ERβ, lamprey ER and amphioxus ER.

In Figure 6 we show the motifs for the ERα and ERβ training set mapped onto a multiple alignment of human ERα, ERβ, lamprey ER and amphioxus ER. Figure 6 also shows the α-helices and β-strands of the 3D structure of ERα and ERβ and amino acids that have been found to be important in binding of estradiol to human ERα [25, 26].

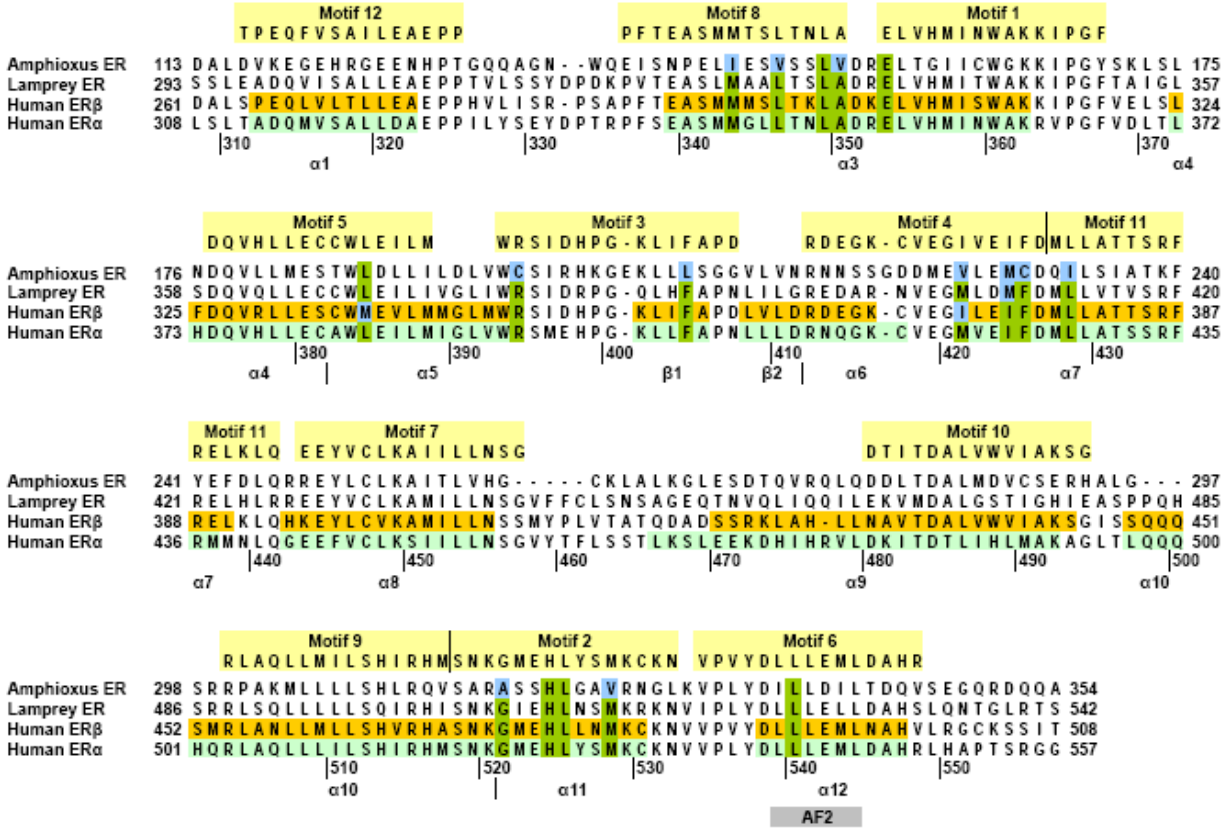

Figure 6. Mapping by MAST of MEME motifs onto sequences of human, lamprey and amphioxus ERs.
The twelve motifs calculated by MEME for the combined ERα and ERβ training set were mapped by MAST onto the sequences of human, lamprey and amphioxus ERs. The α-helices and β-strands from the crystal structures of human ERα and ERβ are shaded. Residues in human ERα that are involved in binding of estradiol [25] are shown in green.

Lamprey lacks motif 10, which maps to α-helix 9 [Figure 6]. Amphioxus ER lacks motifs 2,3,4, 8,10,12 which map to α-helix 11, β-strand 1, α-helices 6 and 7, α-helix 3, α-helix 9

and α-helix 1, respectively. Invertebrate ERs lack motifs 2, 4, 10-12. Oyster and *aplysia* ER contain motif 8; the other invertebrate ERs lack this motif. Amphioxus and human ERR lack motifs 4, 10 and 11.

Table 1 summarizes the order and spacing between motifs ERs and ERRs [Also see Figures 4-6]. Both motif order and spacing are conserved in vertebrate ERs. One intriguing exception is the space of 9 residues between motif 8 and motif 12 in human and chick ERβ. In other vertebrate ERs, this distance is 10 residues. In amphioxus ER, the spacing between motif 7 and motif 9 is only 37 residues, which is close to the distance of 34 residues found in amphioxus and human ERR and several invertebrate ERs. In contrast, vertebrate ERs have a spacing of 45 residues between motifs seven and nine.

Table 1. Motif Order and Spacing for ERα and ERβ Training Set

Human ERβ	[12]-9- [8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-21-[10]-8-[9]-[2]-1-[6]
Chick ERβ	[12]-9- [8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-19-[10]-8-[9]-[2]-1-[6]
X. tropicalis ERβ	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-21-[10]-8-[9]-[2]-1-[6]
Danio ERβ	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-22-[10]-8-[9]-[2]-1-[6]
Cichlid ERβ	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-22-[10]-8-[9]-[2]-1-[6]
Human ERα	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-22-[10]-8-[9]-[2]-1-[6]
Chick ERα	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-22-[10]-8-[9]-[2]-1-[6]
X. tropicalis ERα	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-22-[10]-8-[9]-[2]-1-[6]
Danio ERα	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-22-[10]-8-[9]-[2]-1-[6]
Cichlid ERα	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]-22-[10]-8-[9]-[2]-1-[6]
Lamprey ER	[12]-10-[8]-2-[1]-6-[5]-4-[3]-4-[4]-[11]-1-[7]- 45 -[9]-[2]-1-[6]
Amphioxus ERR	[12]-10-[8]-2-[1]-6-[5]-4-[3]- 34 -[7]- 34 -[9]-[2]-1-[6]
Thais ER	[1]-6-[5]-4-[3]- 35 -[7]- 34 -[9]- 16 -[6]
Human ERR3	[12]-10-[8]-2-[1]-6-[5]-4-[3]- 34 -[7]- 34 -[9]-[2]-1-[6]
Oyster ER	[8]-2-[1]-6-[5]-4-[3]- 35 -[7]- 34 -[9]- 16 -[6]
Amphioxus ER	[1]-6-[5]- 40 -[11]-1-[7]- 37 -[9]- 16 -[6]
Octopus ER	[1]-6-[5]-4-[3]- 35 -[7]- 34 -[9]- 16 -[6]
Aplysia ER	[8]-2-[1]-6-[5]-4-[3]- 35 -[7]- 35 -[9]- 16 -[6]

Phylogenetic analysis of ER and ERR

To better understand the relationship of amphioxus and lamprey ER to other ERs and ERRs, we constructed a phylogenetic tree of their steroid-binding domains. As seen in Figure 7, lamprey ER clusters at the base of the ERα clade, while amphioxus ER on a long branch, indicating that it has diverged substantially from vertebrate ERs, in agreement with the MAST analysis [Figure 4].

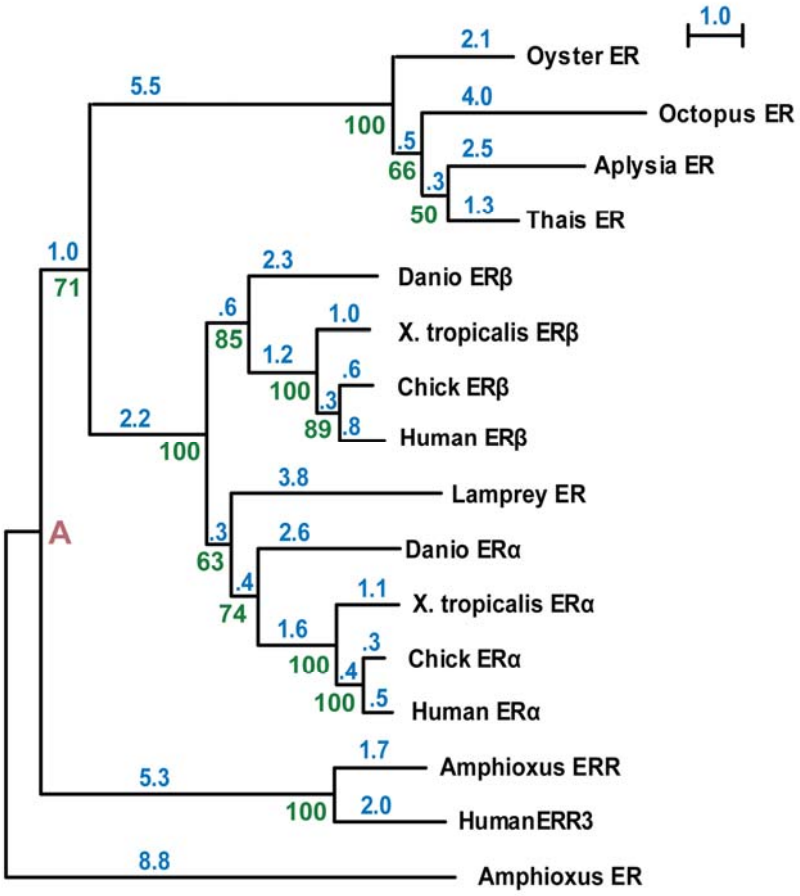

Figure 7. Phylogenetic tree of human ERα, ERβ, ERR, lamprey ER and amphioxus ER and ERR.
Clustal X was used to aligned ERs and ERRs [27]. The phylogenetic tree was constructed with the neighbor-joining algorithm [28] with a correction of branch lengths for rate heterogeneity between sites. Branch lengths are proportional to the distance between proteins. Shown at the nodes are bootstrap values for each branch of the tree, which is the percent this cluster was found in the 1,000 bootstrap trials. Branches with bootstrap values that are greater than fifty percent are significant. Accessions for the sequences are human ERR3 [GenBank:NP_001429], amphioxus ERR [GenBank:AAU88062], *Octopus* ER [GenBank:ABG00286], *Aplysia* ER [GenBank:AAQ95045], *Thais* ER [GenBank:BAC66480] and oyster ER [GenBank:ABI97119].

Discussion

The recently available sequence of an amphioxus ER, when combined with the lamprey ER sequence provides an opportunity to investigate early events in the evolution of the ER. In this report we have subjected these sequences and other ER sequences to motif analysis to investigate similarities and differences among amphioxus ER, lamprey ER, other vertebrate and invertebrate ERs and human and amphioxus ERR.

Lamprey ER

Although steroid binding to lamprey ER and amphioxus ER has not been reported, much is known about the binding of estradiol to mammalian ERs from an analysis of their 3D structures [25, 26], which facilitates analysis of the estrogen-binding domain in lamprey and amphioxus ER. As shown in Figure 6, sixteen out of seventeen amino acids that interact with estradiol on human ERα are identical in lamprey ER. Only Met-409 in lamprey ER differs from Ile-424 in human ERα. MEME and MAST analyses show that lamprey is close to land vertebrate ERs, as found previously [15]. MEME identifies motif 10, which is in the c-terminal half of α-helix 9, as one region in which changes occurred in land and telost ERs during their evolution from their common ancestor with lamprey ER.

Data on the steroid specificity of lamprey ER have not been reported. Interestingly, lamprey has notable differences in its steroid profile in serum compared to humans [29-32]. Lamprey serum contains 15α-hydroxy-steroids, including 15α-hydroxy-estradiol and 15α-hydroxy-estrone. The differences in c-terminal half of lamprey α-helix 9 may be important in binding of 15α-hydroxy-steroids.

Amphioxus ER

In contrast to lamprey ER, amphioxus ER displays significant divergence in key residues that bind estradiol in ERα [25, 26] [Figure 6]. Only six out of seventeen residues are identical. One important difference is Cys-197 in amphioxus ER, instead of Arg-394 in human ERα. As seen in Figure 8, Arg-394 has a stabilizing hydrogen bond with the C3 hydroxyl in estradiol. Moreover, the AR, GR, MR and PR have a corresponding arginine that stabilizes binding of the C3 carbonyl on their cognate steroids [24, 25]. Also, as shown in Figure 8, Phe-404 in human ERα is replaced with Leu-208 in amphioxus ER. Phe-404 stabilizes the side chain of Arg-394 in ERα [25, 26]. A corresponding phenylalanine stabilizes the corresponding arginine in the AR,

GR, MR and PR. Thus, these differences between amphioxus and human ER are likely to be important in steroid binding. Although these and other sequence differences suggest that amphioxus ER binds novel steroids, there is a caveat that the tertiary structure of amphioxus ER may fold so that other residues stabilize estradiol.

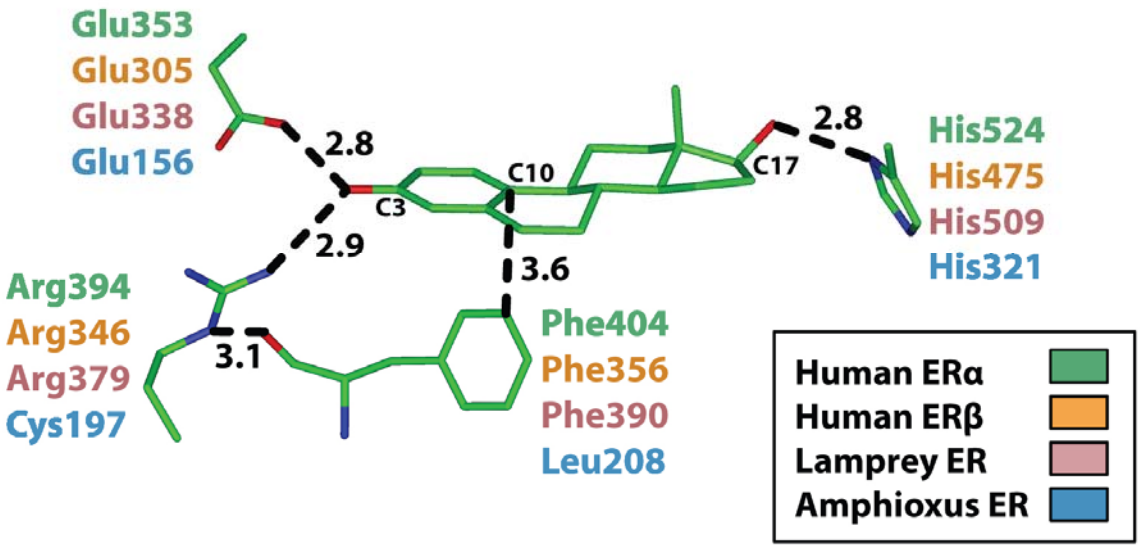

Figure 8. Comparison of key estrogen-binding residues in human ERα with corresponding residues in lamprey and amphioxus ER.
Amino acids in the estrogen-binding site of human ERα [25] are shown along with those residues in human ERβ, amphioxus and lamprey ER that align in Figure 6. Glu-353, Arg-394, Phe-404 and His-524 are conserved in vertebrate ERs. Amphioxus ER Glu-156 and His-321 match Glu-353 and His 524. However, amphioxus ER Cys-197 and Leu-208 match Arg-394 and Phe-404, respectively. Lamprey ER conserves all four residues.

Human ERα and ERβ

A nice surprise was that MEME uncovered differences between the estrogen-binding domains on human ERα and ERβ. As seen in Figures 3 to 5, there are differences between human ERα and ERβ in the location of several motifs, even though human ERα and ERβ sequences are 59% identical [83% positive matches]. This sequence similarity is reflected in similar, but not identical, affinities for estrogens and various plant-derived chemicals such as genistein, a component in soy [33]. In fact, differences in ligand-binding between human ERα and ERβ are of interest because in some tissues these ERs appear to have opposite activities [34-

36]. For example, ERα promotes prostate cell proliferation, while ERβ inhibits proliferation. Thus, there is considerable effort to find chemicals that specifically stimulate ERβ-mediated transcription or inhibit ERα-mediated transcription. Motif analysis of human ERα and ERβ may help in structural analyses in pursuit of selective drugs to modulate the actions of these ERs.

Evolution of Steroid Binding to the ER

Because steroid receptors are absent in Ciona [8], amphioxus is the most basal deuterostome in which there appears to be an ER that could be activated by steroids [16, 37, 38]. The phylogenetic tree shows that the steroid-binding domain on amphioxus ER is 8.8 units from node A, from where ERRs and vertebrate and invertebrate ERs diverge [Figure 7]. In contrast, amphioxus ERR and human ERR are 7 units and 7.3 units, respectively, from node A. *Thais* ER and octopus ER are 7.6 units and 11 units, respectively, from node A. Thus, the phylogenetic analysis of ERs and ERRs is in agreement with the MAST scores, which have amphioxus ER distant from vertebrate ERs. MAST analysis also places amphioxus ERR closer to vertebrate ERs than is either amphioxus ER or invertebrate ERs.

The long distance of amphioxus ER branch from node A suggests that this ER has undergone substantial changes in sequence after it diverged from the line leading to vertebrates. A rapidly evolving amphioxus ER sequence may be due to specific responses of *B. floridae* to natural selection or, alternatively, extensive sequence changes may be found in other amphioxus ERs. Regarding the first possibility, significant differences in the rate of sequence change have been found between two roundworm species, *Caenorhabditis* and *Trichinella* [39], and two tunicates, *Ciona intestinalis* and *Oikopleura dioica* [40]. In the light of these two examples of uneven rate of sequence evolution in related species, it may be that there has been less sequence change in ER from *B. belcheri* or another amphioxus species during their divergence from a common ancestor with vertebrate ERs. If such amphioxus ERs exist, then their sequences would assist in elucidating the early stages in the evolution of estrogen binding in nuclear receptors.

It is not known which ligands regulate transcriptional activity of amphioxus ER [35]. There is evidence that amphioxus can synthesize estradiol [16, 38], but the presence of various estradiol derivatives, such as 15α-hydroxy-estradiol, has not been investigated. The ligand-binding domain in amphioxus ER and human ERα are about 35% identical, which suggests differences in substrate specificity, especially in view of the low conservation in *B. floridae* ER of key residues that bind estradiol in human ERα [Figures 6, 8] and differences in conservation

of motifs [Figure 4]. It may be that *B. floridae* ER is regulated by a novel estradiol derivative, such as 15α-hydroxy-estradiol [29-32], or other steroids, such as a Δ5 steroid [17, 18, 33], or possibly more novel ligands.

Conclusions.

Motif analysis identifies regions of similarity and divergence between amphioxus ER and vertebrate ERs, as well as similarities and differences between vertebrate ERs and chordate ERRs and invertebrate ERs. There are significant differences between *B. floridae* ER and vertebrate ERs in the steroid-binding domain as measured by conservation of motifs and percent of amino acids in ERα that are known to stabilize estradiol. Novel steroid(s) may regulate transcriptional activity of *B. floridae* ER. Sequences of ERs from other amphioxus species, such as *B. belcheri*, are needed to better understand the evolution of structure and function in estrogen-binding domain in vertebrate ERs. The absence in lamprey ER of motif 10, which maps to the c-terminus half of α-helix 9, may be important in recognition of estrogens, such as 15α-hydroxy-estradiol.

Methods

Motif analysis

Motifs for vertebrate ERs were discovered using MEME [**M**ultiple **E**xpectation-maximum for **M**otif **E**licitation], which has been described in detail elsewhere [19, 20]. Briefly, MEME is an artificial-intelligence-based motif analysis tool that, given a set of unaligned sequences, identifies in an unbiased, automated fashion the conserved regions [i.e. motifs] that are characteristic of the dataset [e.g. in this paper, vertebrate ERs]. MEME uses statistical modeling techniques to automatically choose the best width, number of occurrences, and description for each motif in a collection of ER sequences. The MEME output includes a representation of each motif in the input dataset [e.g. vertebrate ERs] as a position-dependent probability matrix, or log-odds matrix. Each column of the matrix gives the probabilities of each residue at that position.

The log-odds matrix for each motif can be used as input into MAST [**M**otif **A**lignment and **S**earch **T**ool], which can search databases such as Genbank or as reported here a database consisting of invertebrate and vertebrate ERs and ERRs. We used MAST to calculate match

scores for various these ERs and ERRs and a schematic diagram showing the order and spacing of motifs within each ER or ERR sequence.

Selection of the ER training set for MEME analysis

In collecting sequences for MEME, we were limited by the strong sequence conservation of vertebrate ERα and ERβ. The amino acid sequences of human ERα and mouse ERα are 95% identical, with 97% positive matches. Positive matches include conservative replacements such as lysine and arginine or aspartic acid and glutamic acid. Thus, there is little additional information about the divergence of an estrogen receptor, when comparing human and mouse ERα. Comparisons of human ERα and ERβ with non-mammalian ERs were more encouraging. For example, human and *Xenopus tropicalis* ERα are 80% identical, with 92% positive matches, and human and *X. tropicalis* ERβ are 75% identical, with 86% positive matches. Human ERα and ERβ are 59% identical, with 83% positive matches. With this sequence conservation in mind, we selected ER sequences from human, chicken, frog, and two fish: zebrafish and cichlid for MEME analysis, to provide MEME with a diverse training set for investigating differences and similarities between amphioxus ER and vertebrate ERs.

The training set of vertebrate ERs consisted of human ERα [SwissProt:P03372] and ERβ [SwissProt:Q92731], chicken ERα [SwissProt:P06212] and ERβ [SwissProt:Q9PTU5], *Xenopus tropicalis* ERα [GenBank:NP_988866] and ERβ [GenBank:NP_001035101], *Danio rerio* ERα [SwissProt:P57717] and ERβ [GenBank:AAK16742], and cichlid ERα [GenBank:AAR82891] and ERβ [GenBank:ABI18967]. MEME calculated the first 12 motifs of the training set of ERα and ERβ, selecting optimal motifs between six and fifteen amino acids. We limited the motif length to fifteen amino acids to increase the resolution of the analysis of ERs and ERRs. With the exception of motif eleven, which contained fourteen residues, the other motifs contained fifteen residues. Thus, the motifs characterize 179 residues of the training set. MEME also calculated 12 motifs for the ERα training set and the ERβ training set. Then MAST mapped the twelve motifs onto various ERs and ERRs.

Clustal X [27] was used to construct a multiple alignment of ERs and ERRs for construction of a phylogenetic tree using the neighbor-joining algorithm [28] with a correction of branch lengths for rate heterogeneity between sites.

Authors Contributions

MEB conceived of this project, supervised the research and drafted the manuscript. CC carried out motif and phylogenetic analysis and preparation of the figures.

References

1. Baker ME: **Steroid receptor phylogeny and vertebrate origins**. *Mol Cell Endocrinol* 1997, **135**(2):101-107.
2. Escriva H, Delaunay F, Laudet V: **Ligand binding and nuclear receptor evolution**. *Bioessays* 2000, **22**(8):717-727.
3. Olefsky JM: **Nuclear receptor minireview series**. *J Biol Chem* 2001, **276**(40):36863-36864.
4. Bertrand S, Brunet FG, Escriva H, Parmentier G, Laudet V, Robinson-Rechavi M: **Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems**. *Mol Biol Evol* 2004, **21**(10):1923-1937.
5. Baker ME: **Xenobiotics and the evolution of multicellular animals: emergence and diversification of ligand-activated transcription factors**. *Integrative and Comparative Biology* 2005, **45**:172-178.
6. Baker ME: **Evolution of adrenal and sex steroid action in vertebrates: a ligand-based mechanism for complexity**. *Bioessays* 2003, **25**(4):396-400.
7. Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH: **Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus**. *Dev Biol* 2006, **300**(1):90-107.
8. Campbell RK, Satoh N, Degnan BM: **Piecing together evolution of the vertebrate endocrine system**. *Trends Genet* 2004, **20**(8):359-366.
9. Thornton JW, Need E, Crews D: **Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling**. *Science* 2003, **301**(5640):1714-1717.
10. Keay J, Bridgham JT, Thornton JW: **The Octopus vulgaris estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications**. *Endocrinology* 2006, **147**(8):3861-3869.
11. Kajiwara M, Kuraku S, Kurokawa T, Kato K, Toda S, Hirose H, Takahashi S, Shibata Y, Iguchi T, Matsumoto T *et al*: **Tissue preferential expression of estrogen receptor gene in the marine snail, Thais clavigera**. *Gen Comp Endocrinol* 2006, **148**(3):315-326.
12. Bannister R, Beresford N, May D, Routledge EJ, Jobling S, Rand-Weaver M: **Novel estrogen receptor-related Transcripts in Marisa cornuarietis; a freshwater snail with reported sensitivity to estrogenic chemicals**. *Environ Sci Technol* 2007, **41**(7):2643-2650.
13. Matsumoto T, Nakamura AM, Mori K, Akiyama I, Hirose H, Takahashi Y: **Oyster estrogen receptor: cDNA cloning and immunolocalization**. *Gen Comp Endocrinol* 2007, **151**(2):195-201.
14. Bridgham JT, Carroll SM, Thornton JW: **Evolution of hormone-receptor complexity by molecular exploitation**. *Science* 2006, **312**(5770):97-101.
15. Thornton JW: **Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions**. *Proc Natl Acad Sci U S A* 2001, **98**(10):5671-5676.

16. Mizuta T, Kubokawa K: **Presence of sex steroids and cytochrome P450 genes in amphioxus**. *Endocrinology* 2007, **148**(8):3554-3565.
17. Baker ME: **Recent insights into the origins of adrenal and sex steroid receptors**. *J Mol Endocrinol* 2002, **28**(3):149-152.
18. Baker ME: **Co-evolution of steroidogenic and steroid-inactivating enzymes and adrenal and sex steroid receptors**. *Mol Cell Endocrinol* 2004, **215**(1-2):55-62.
19. Bailey TL, Baker ME, Elkan CP: **An artificial intelligence approach to motif discovery in protein sequences: application to steriod dehydrogenases**. *J Steroid Biochem Mol Biol* 1997, **62**(1):29-44.
20. Grundy WN, Bailey TL, Elkan CP, Baker ME: **Hidden Markov model analysis of motifs in steroid dehydrogenases and their homologs**. *Biochem Biophys Res Commun* 1997, **231**(3):760-766.
21. Giguere V: **To ERR in the estrogen pathway**. *Trends Endocrinol Metab* 2002, **13**(5):220-225.
22. Greschik H, Wurtz JM, Sanglier S, Bourguet W, van Dorsselaer A, Moras D, Renaud JP: **Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3**. *Mol Cell* 2002, **9**(2):303-313.
23. Kallen J, Schlaeppi JM, Bitsch F, Filipuzzi I, Schilb A, Riou V, Graham A, Strauss A, Geiser M, Fournier B: **Evidence for ligand-independent transcriptional activation of the human estrogen-related receptor alpha (ERRalpha): crystal structure of ERRalpha ligand binding domain in complex with peroxisome proliferator-activated receptor coactivator-1alpha**. *J Biol Chem* 2004, **279**(47):49330-49337.
24. Baker ME, Chandsawangbhuwana C: **Analysis of 3D models of octopus estrogen receptor with estradiol: evidence for steric clashes that prevent estrogen binding**. *Biochem Biophys Res Commun* 2007, **361**(3):782-788.
25. Tanenbaum DM, Wang Y, Williams SP, Sigler PB: **Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains**. *Proc Natl Acad Sci U S A* 1998, **95**(11):5998-6003.
26. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M: **Molecular basis of agonism and antagonism in the oestrogen receptor**. *Nature* 1997, **389**(6652):753-758.
27. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools**. *Nucleic Acids Res* 1997, **25**(24):4876-4882.
28. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**(4):406-425.
29. Lowartz SM, Renaud RL, Beamish FW, Leatherland JF: **Evidence for 15alpha- and 7alpha-hydroxylase activity in gonadal tissue of the early-life stages of sea lampreys, Petromyzon marinus**. *Comp Biochem Physiol B Biochem Mol Biol* 2004, **138**(2):119-127.
30. Lowartz S, Petkam R, Renaud R, Beamish FW, Kime DE, Raeside J, Leatherland JF: **Blood steroid profile and in vitro steroidogenesis by ovarian follicles and testis fragments of adult sea lamprey, Petromyzon marinus**. *Comp Biochem Physiol A Mol Integr Physiol* 2003, **134**(2):365-376.
31. Bryan MB, Scott AP, Cerny I, Young BA, Li W: **15Alpha-hydroxyprogesterone in male sea lampreys, Petromyzon marinus L**. *Steroids* 2004, **69**(7):473-481.

32. Bryan MB, Young BA, Close DA, Semeyn J, Robinson TC, Bayer J, Li W: **Comparison of synthesis of 15 alpha-hydroxylated steroids in males of four North American lamprey species**. *Gen Comp Endocrinol* 2006, **146**(2):149-156.
33. Kuiper GG, Carlsson B, Grandien K, Enmark E, Haggblad J, Nilsson S, Gustafsson JA: **Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta**. *Endocrinology* 1997, **138**(3):863-870.
34. Weihua Z, Andersson S, Cheng G, Simpson ER, Warner M, Gustafsson JA: **Update on estrogen signaling**. *FEBS Lett* 2003, **546**(1):17-24.
35. Weihua Z, Lathe R, Warner M, Gustafsson JA: **An endocrine pathway in the prostate, ERbeta, AR, 5alpha-androstane-3beta,17beta-diol, and CYP7B1, regulates prostate growth**. *Proc Natl Acad Sci U S A* 2002, **99**(21):13589-13594.
36. Deroo BJ, Korach KS: **Estrogen receptors and human disease**. *J Clin Invest* 2006, **116**(3):561-570.
37. Baker ME: **Amphioxus, a primitive chordate, is on steroids: evidence for sex steroids and steroidogenic enzymes**. *Endocrinology* 2007, **148**(8):3551-3553.
38. Callard GV, Pudney JA, Kendall SL, Reinboth R: **In vitro conversion of androgen to estrogen in amphioxus gonadal tissues**. *Gen Comp Endocrinol* 1984, **56**(1):53-58.
39. Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA: **Evidence for a clade of nematodes, arthropods and other moulting animals**. *Nature* 1997, **387**(6632):489-493.
40. Delsuc F, Brinkmann H, Chourrout D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates**. *Nature* 2006, **439**(7079):965-968.